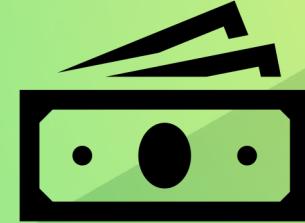


BUY NOW, PAY LATER



GROUP 10 -

Arshnoor Kaur, Hannah Chin, Kasturi Deshpande, Aishwarya Hirve, Aoi Fujii



TEAM



Arshnoor Kaur



Hannah Chin



Kasturi Deshpande



Aishwarya Hirve



Aoi Fujii



AGENDA

- ▶ Introduction and Overview
- ▶ Preliminary Analysis
- ▶ Feature Generation
- ▶ Ranking model
- ▶ Insights
- ▶ Recommendations
- ▶ Limitations
- ▶ Privacy Compliance

INTRODUCTION AND PROJECT BRIEF

Project brief



Pay in 5 installments feature



Merchants are seeking to form partnership with BNPL to boost customer base

BNPL is looking to partner with 100 merchants to pilot “Pay in 5 Installments” feature

Our Aim



Determine the Top 100 merchants
to partner with the BNPL firm

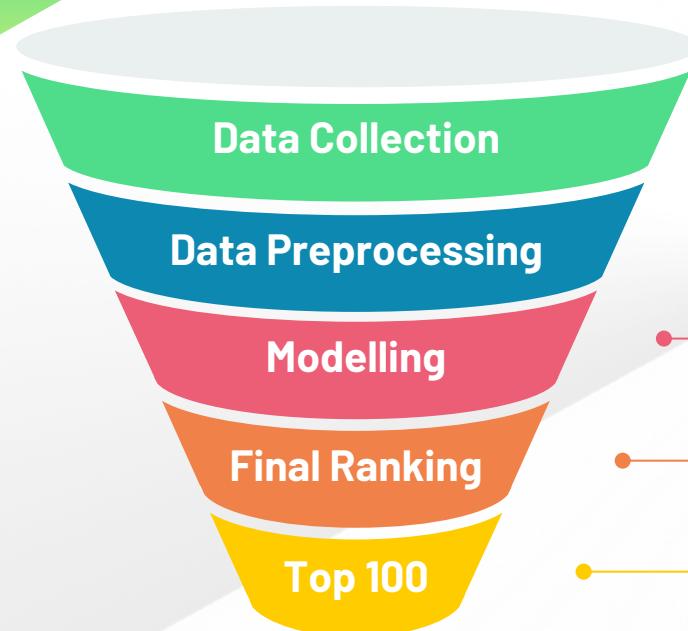


Establish distinguishing features of
merchants to enable successful partnership

PROJECT OVERVIEW AND TIMELINE



RANKING PROCESS OVERVIEW



- Collect data to investigate merchants
- Prepare datasets for modelling
- Create models to predict merchants' future insights
- Final model to rank each merchant
Based on the predicted insights
- Final Ranking of merchants



TIMELINE

- 29th Aug – 4th Sept
- Data extraction and preliminary analysis (10 hours)

- 12th Sept – 18th Sept
- Curation of dataset and analysis of fraud (12 hours)

- 26th Sept – 2nd Oct
- Creation of ranking model (8 hours)



- 5th Sept – 11th Sept
- Outlier analysis and insights (7 hours)

- 19th Sept – 25th Sep
- Creation of feature models and merchant categorisation (17 hours)

- 3rd Oct – 10th Oct
- Final ranking of merchants (10 hours)



COLLABORATION



GitHub



Trello



Confluence



Zoom

DATASETS AND PREPROCESSING





DATASETS

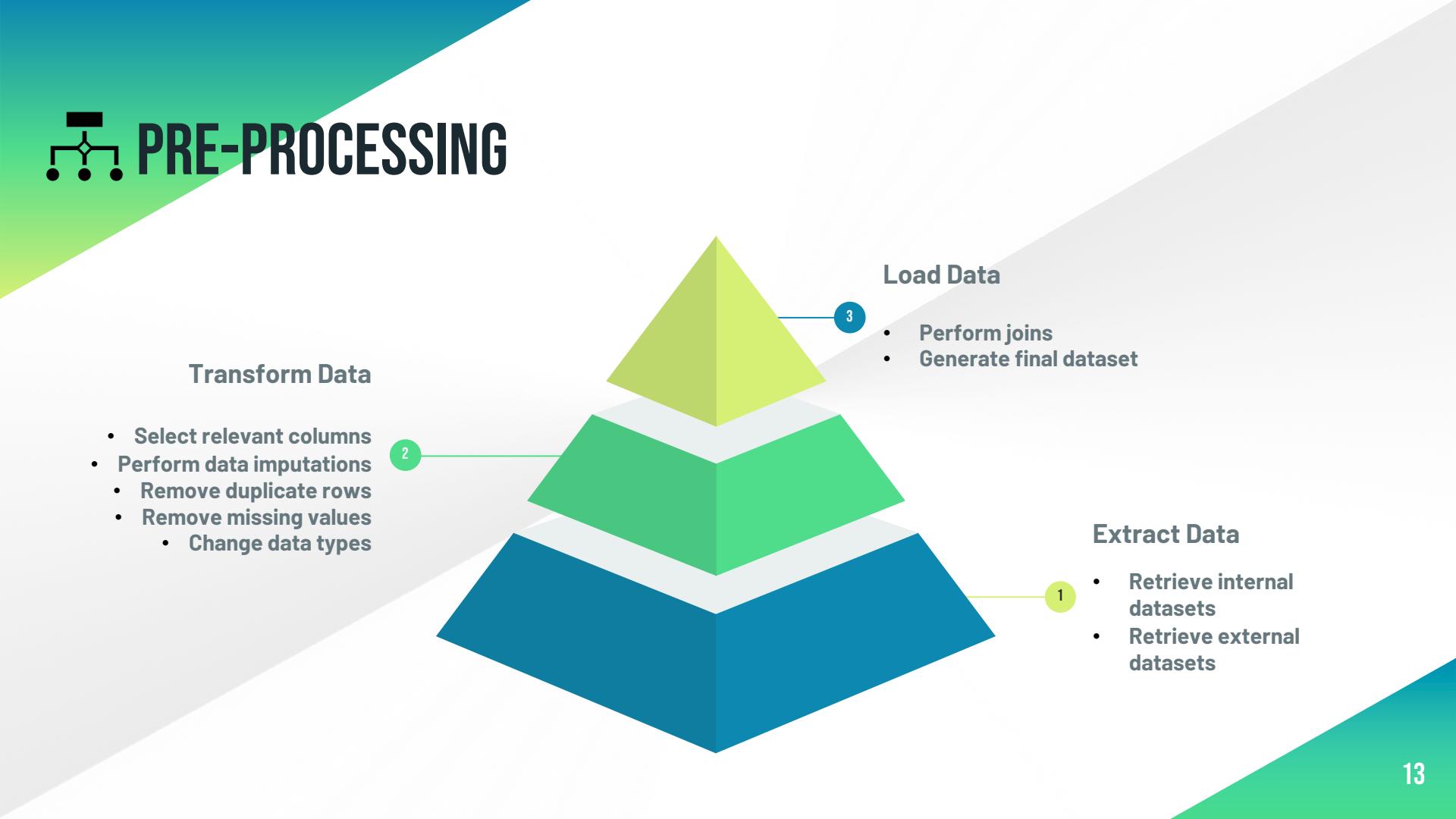
Internal

- Merchant data
- Customer data
- Transactions data
- Fraud probabilities for merchant and customers



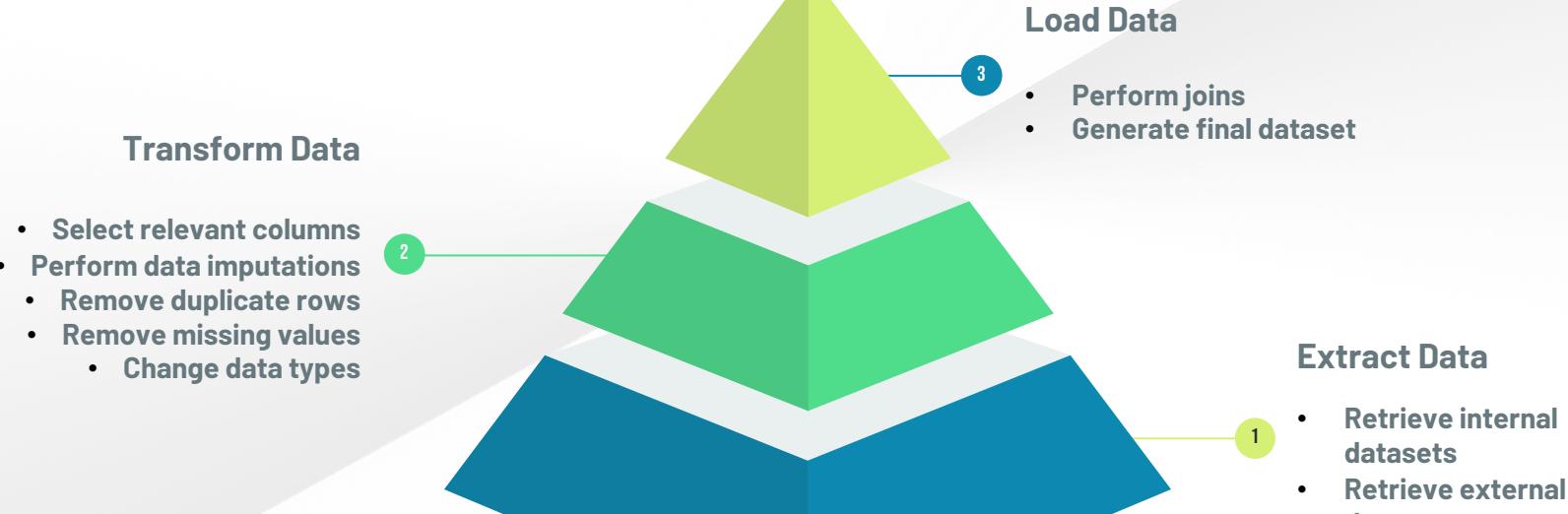
External

- Income data by SA2 areas
- Census data by SA2 areas
- Total population by SA2 areas
- Covid-19 data
- Geospatial data



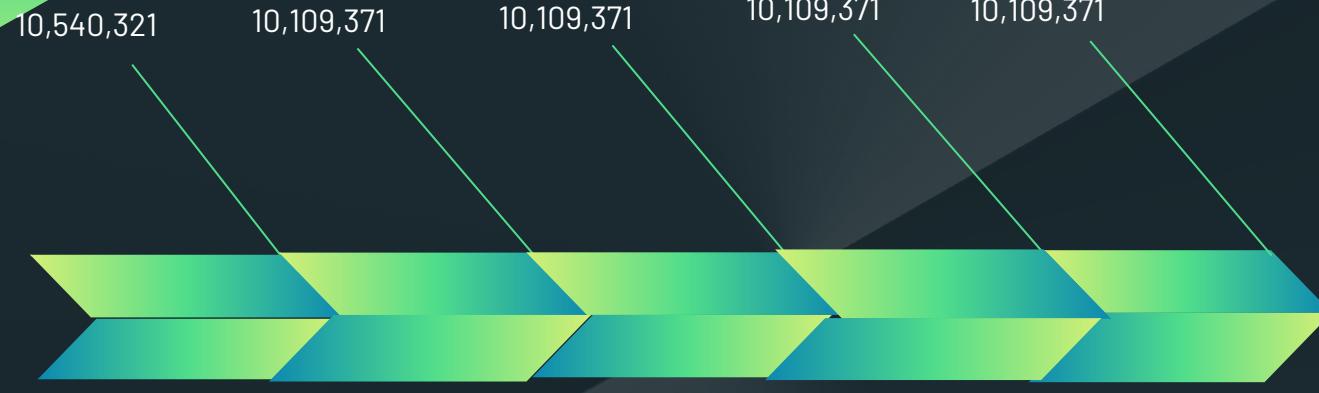
PRE-PROCESSING

- ## Transform Data
- Select relevant columns
 - Perform data imputations
 - Remove duplicate rows
 - Remove missing values
 - Change data types



- ## Extract Data
- Retrieve internal datasets
 - Retrieve external datasets

OUTLIER ANALYSIS



Original count

Internal +
external data
together

Valid Merchant ABN

Remove all the
instances where
merchant Abn is
Null

Transactions with non \$0

Remove all the
instances which
have a negative
dollar value or a
value equal to 0

Valid Customer ID

Remove all
the instances
with customer
ID equal to
null

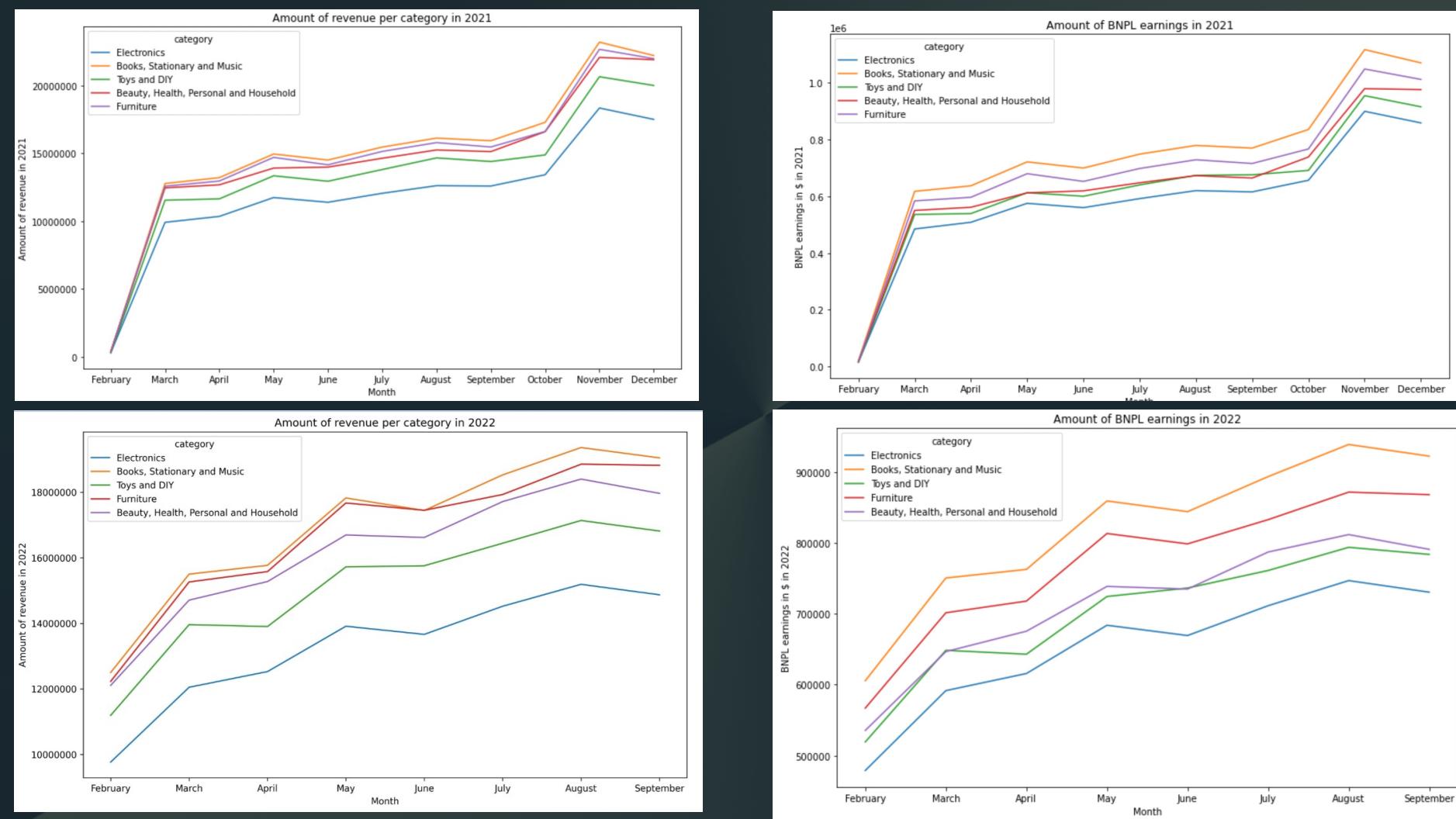
Valid gender values

Remove all the
instances with null
values for gender

Preliminary Analysis

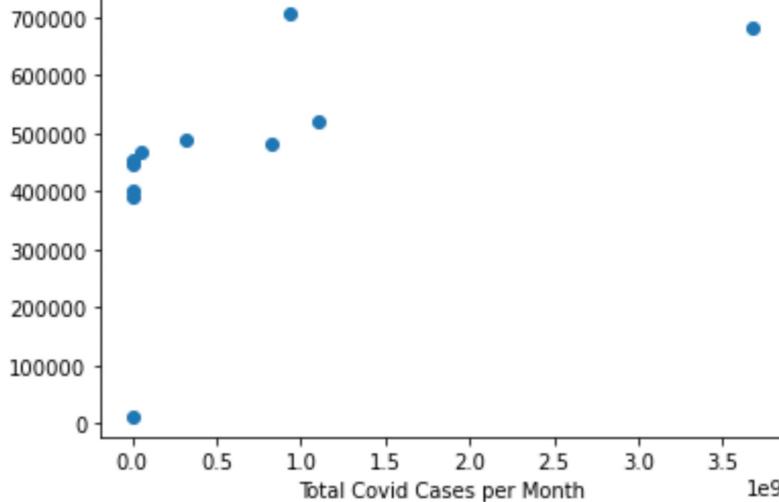
DATA EXPLORATION





Covid Cases per Month vs Number of Transactions per Month

Number of Transactions per Month

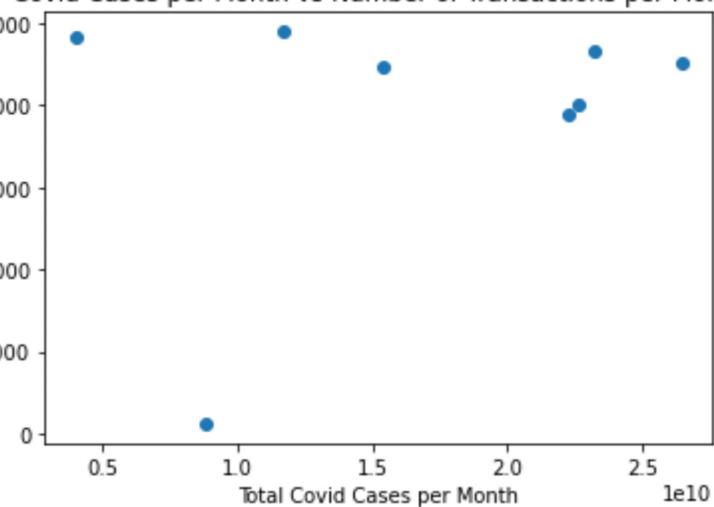


2021

2022

Covid Cases per Month vs Number of Transactions per Month

Number of Transactions per Month



Total Covid Cases per Month

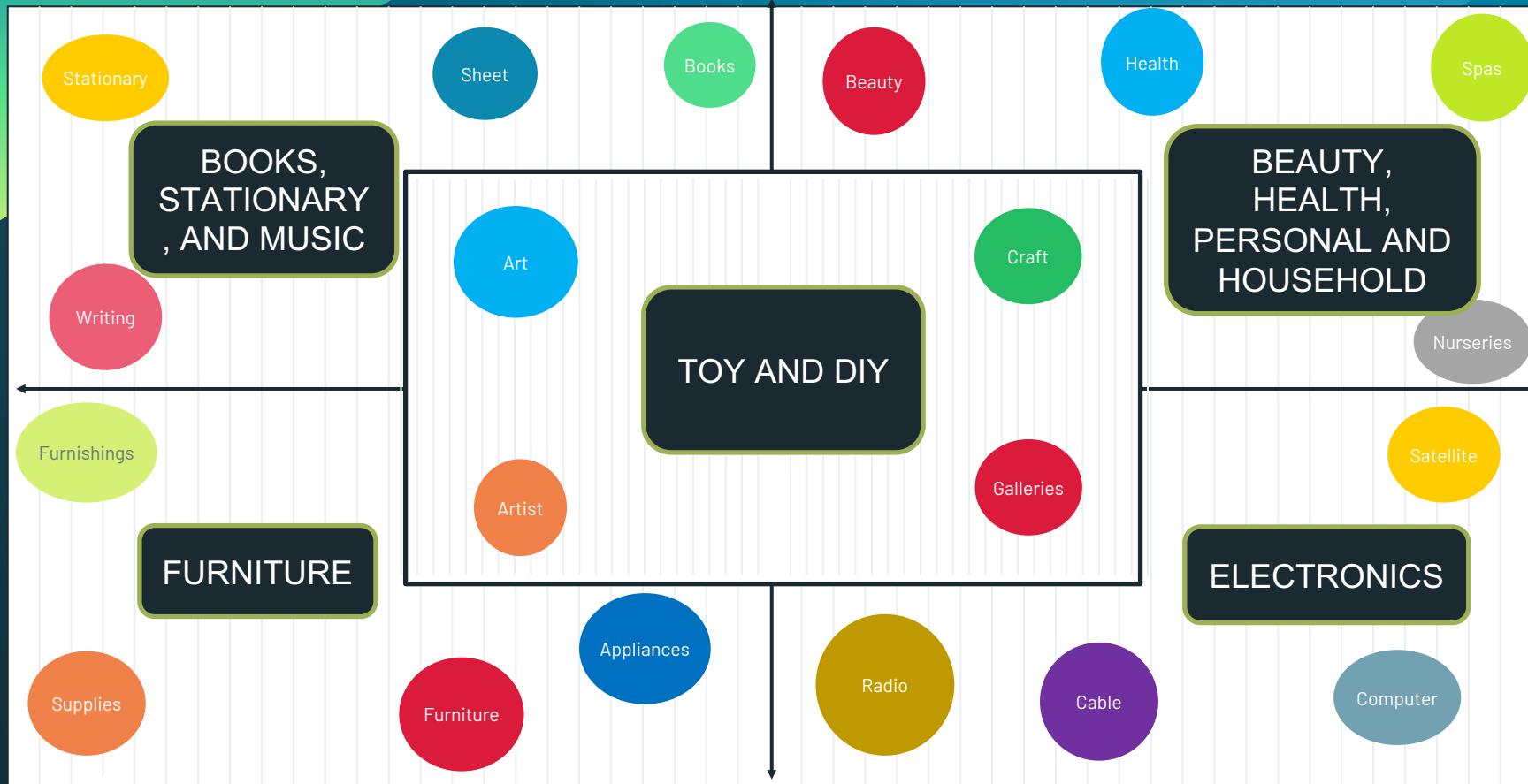
1e10

Feature generation

MERCHANT CATEGORIZATION



LDA – Latent Dirichlet Allocation



FRAUD FEATURE





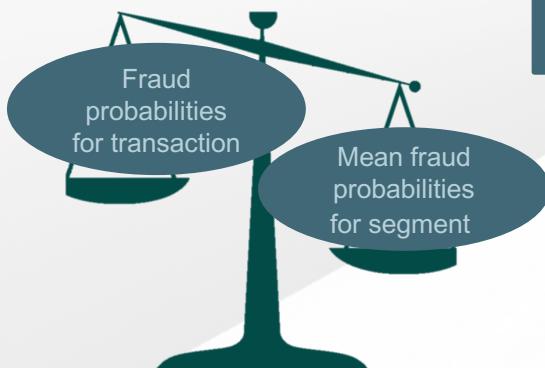
FRAUD PROBABILITY CALCULATION



Calculate mean consumer and merchant fraud probabilities for each segments



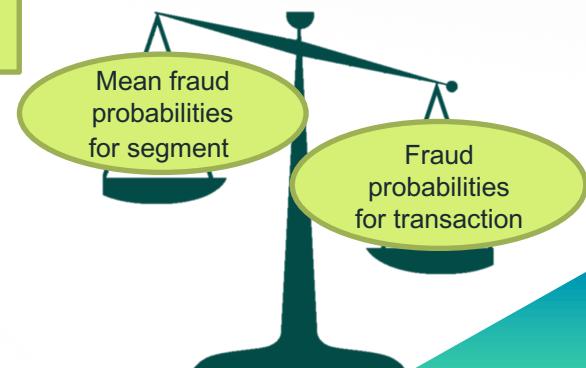
Compare fraud probabilities of transaction with mean fraud probabilities of relevant segment



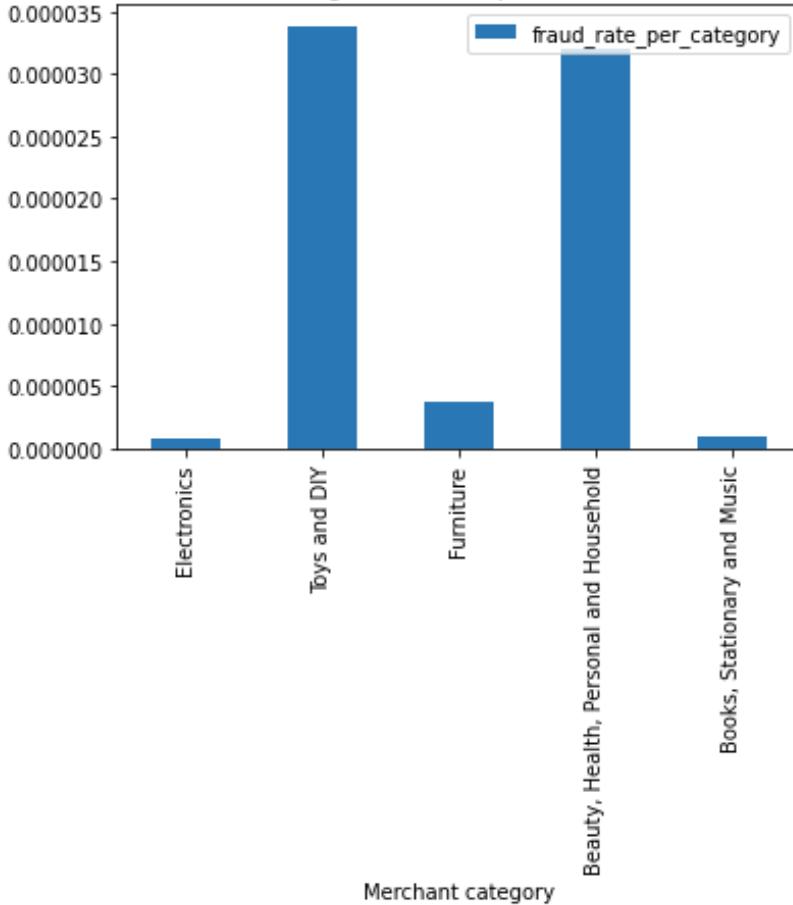
Mark transaction
as fraudulent

Mark transaction
as genuine

Calculate the mean number of
fraudulent transactions per
merchant

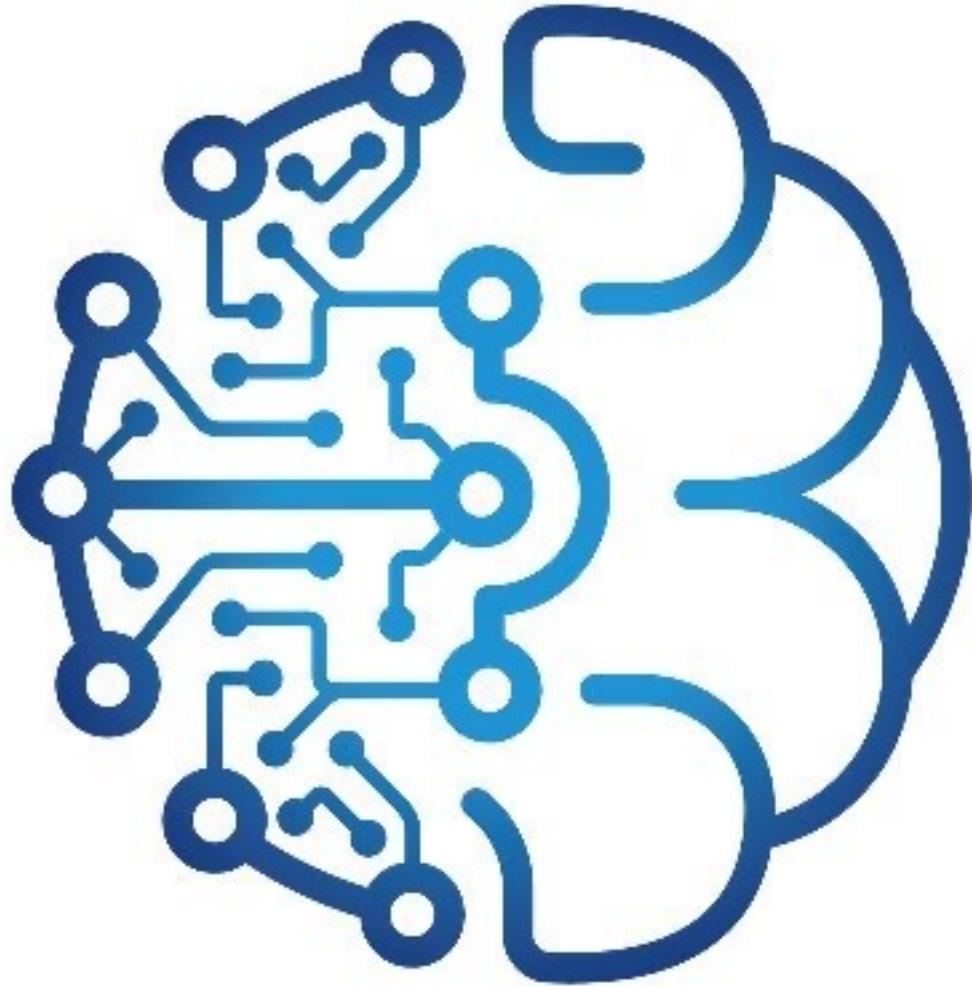


Average fraud rate per Merchant



Feature generation

MACHINE LEARNING MODELS



FEATURE GENERATION USING ML TO BUILD A RANKING MODEL



Revenue

No. of Customers

No. of Transactions

BNPL firm earnings

Prediction type

Predicted revenue for the next month for each merchant

Predicted no. of customers in the next month

Predicted no. of transactions for the next month

Predicted share of earnings for the BNPL firm in the next month

Mean Absolute Error

209.456481

2.093303

2.695082

10.325495

Important Features

- Total earnings
- No. of transactions
- Category – Beauty, Health, Personal and Household

- No. of customers
- Total earnings
- Income per person

- No. of transactions
- Total earnings
- Income per person

- BNPL earnings
- No. of transactions
- Take rate

FINAL RANKING MODEL

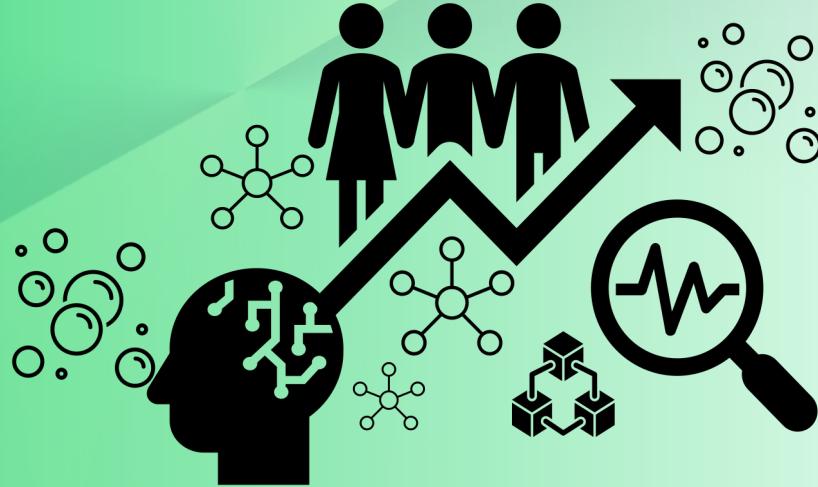
Combining the five features to rank our merchants according to the following weights -

1. Projected BNPL Earnings
2. Probability of Fraud
3. Projected Number of Unique Customers
4. Projected Number of Transactions
5. Projected Merchant Revenue



Insights

INSIGHTS



BOOKS, STATIONARY, AND MUSIC TURNS OUT TO BE THE MOST DESIRABLE CATEGORY TO PARTNER WITH

Top Category

BNPL earnings

No. of Transactions

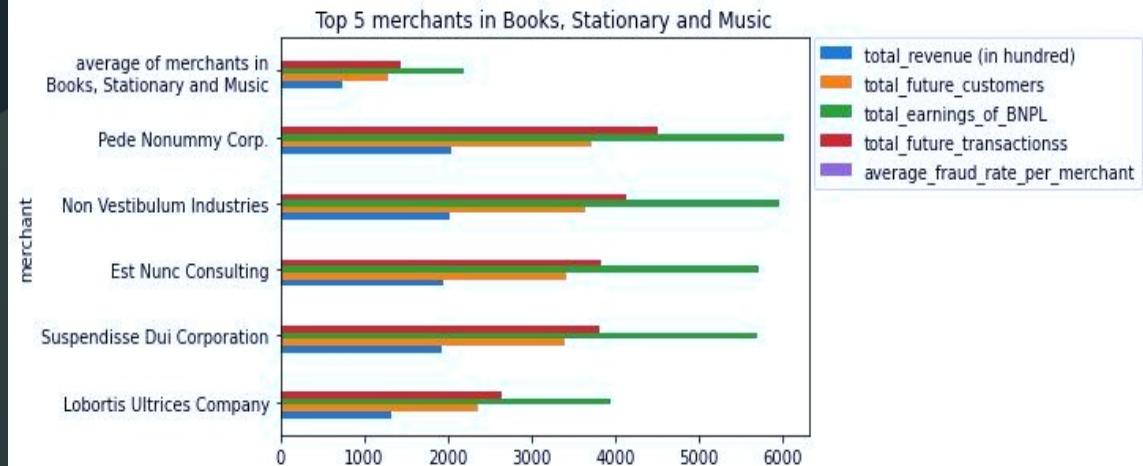
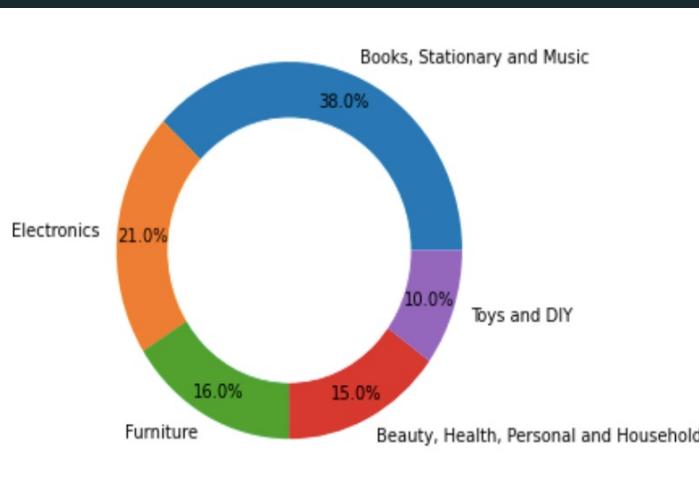
No. of Customers

Fraud probability

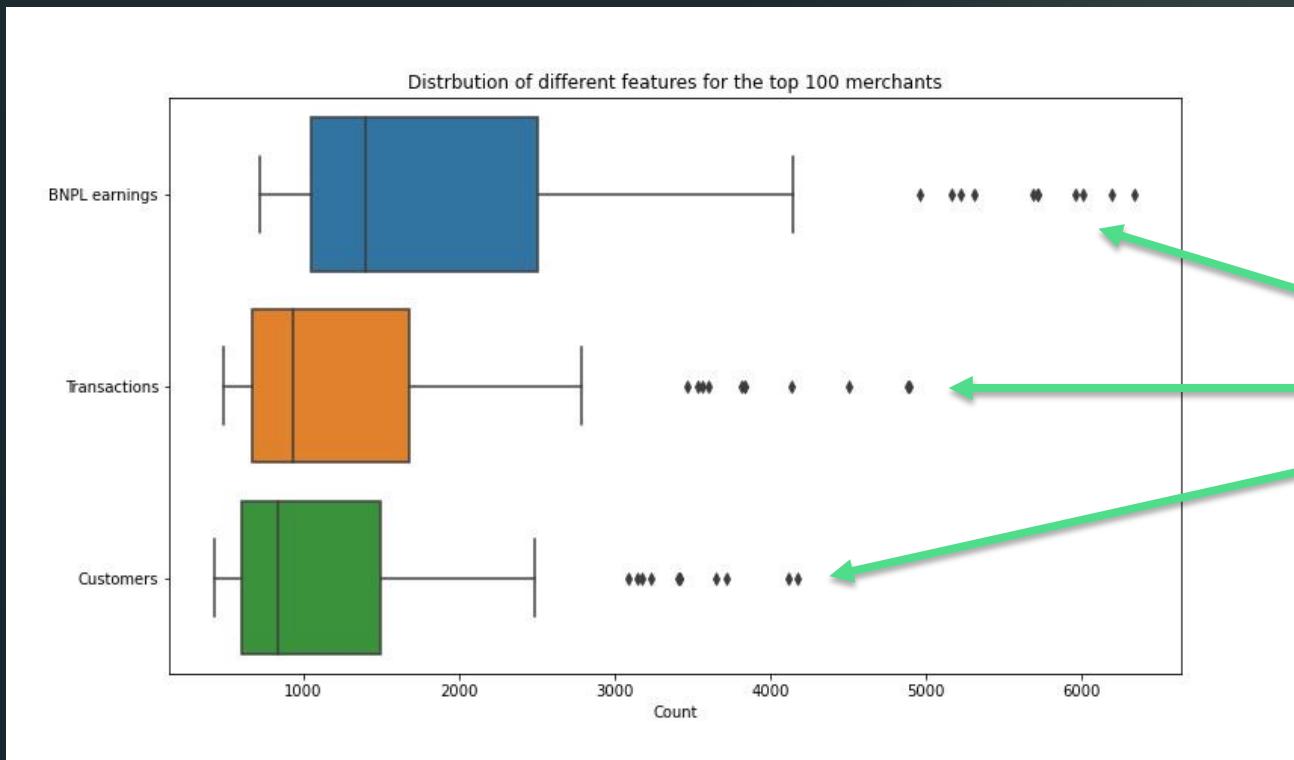
Revenue

38

Books, Stationary,
and Music



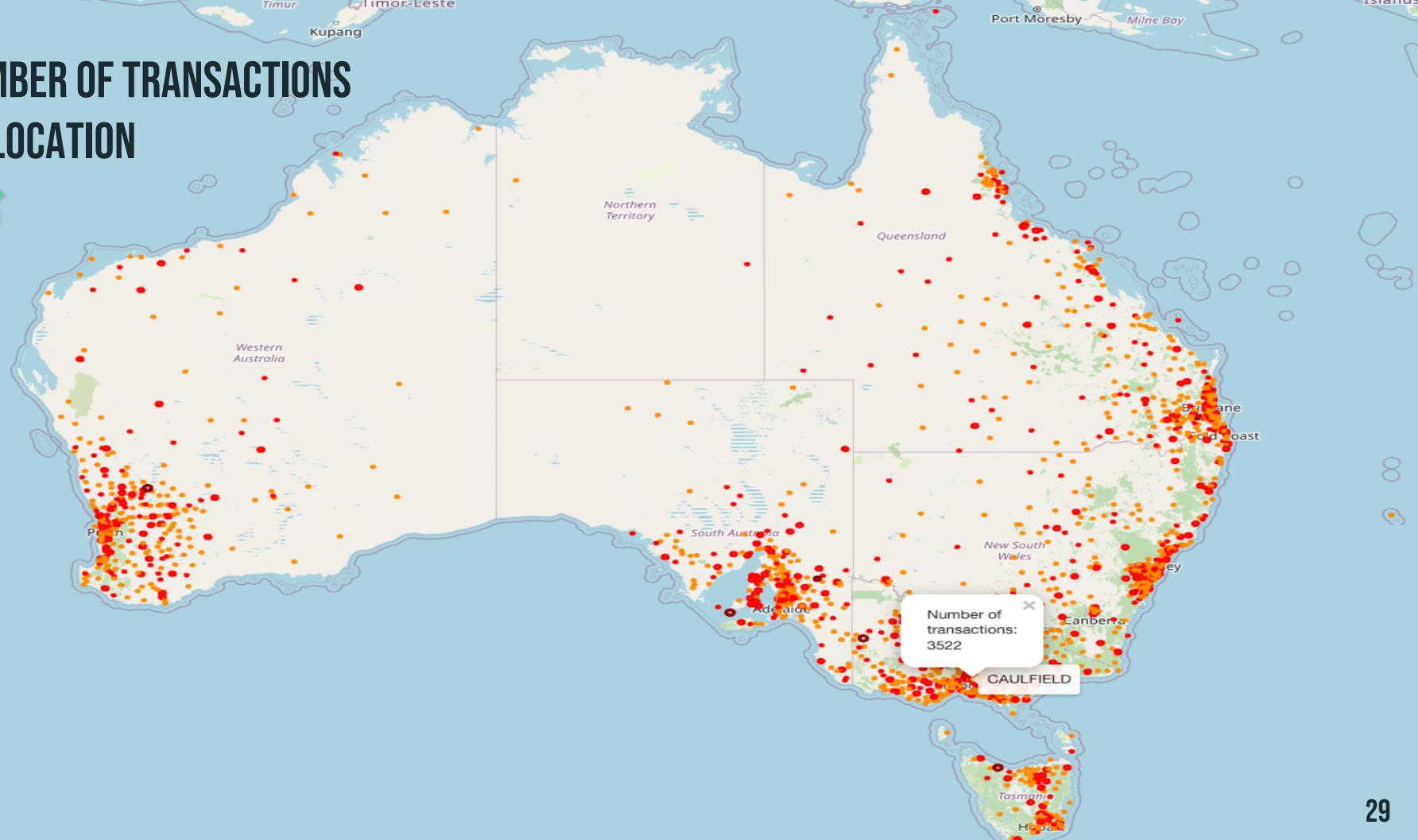
THE TOP 10 MERCHANTS IN THE TOP 100 PERFORM SIGNIFICANTLY BETTER



Irrespective of their category, the BNPL firm can partner with these merchants as they show high performance

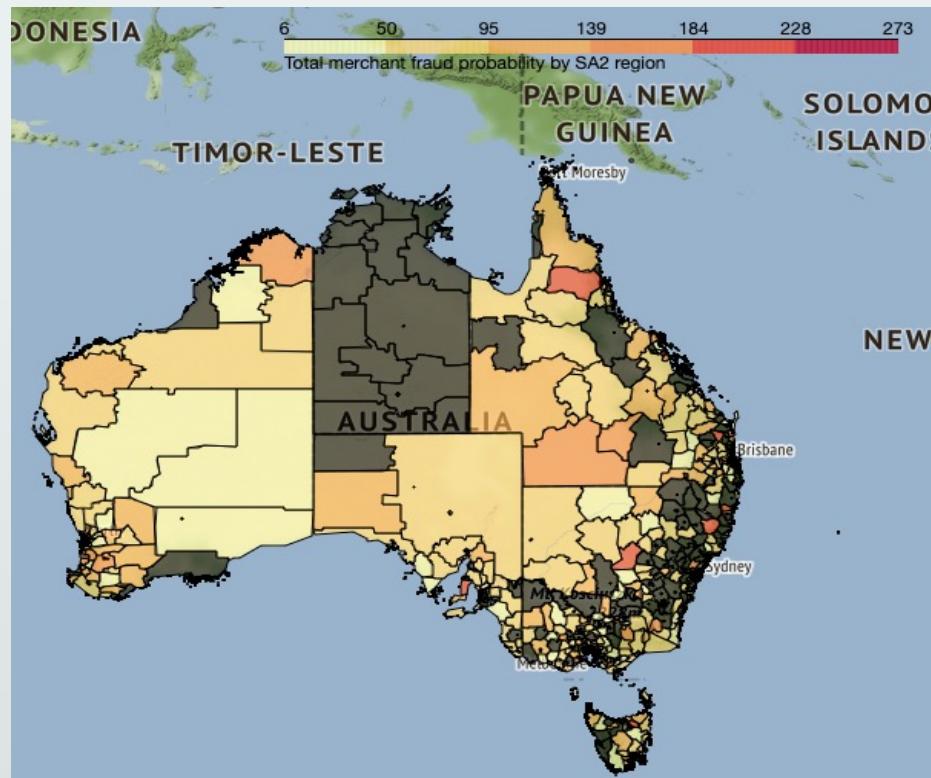
Exceptional performers

NUMBER OF TRANSACTIONS BY LOCATION

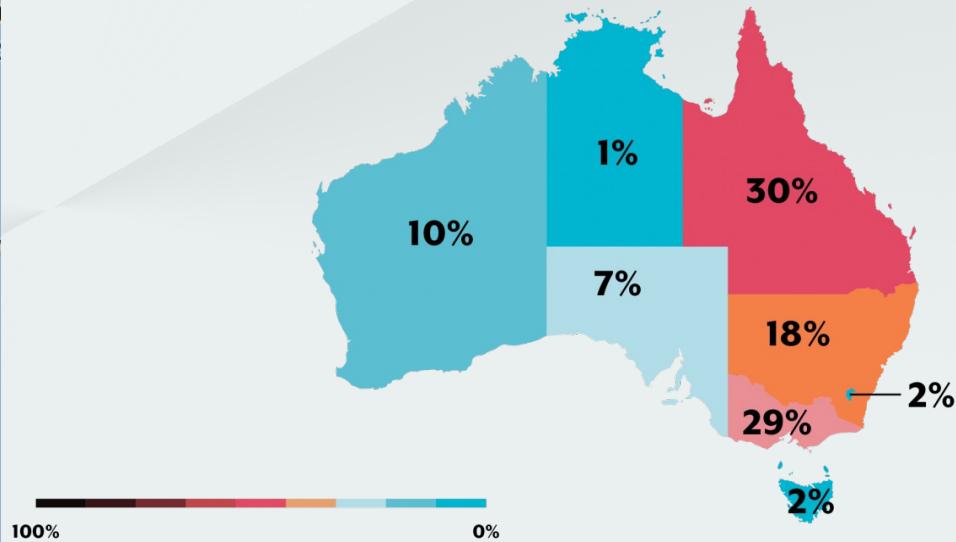




MERCHANT FRAUD BY LOCATION



CYBER CRIME INCIDENTS BY LOCATION





ASSUMPTIONS

- ▶ Probability distributions of simulated transactions mimic those of real transactions
- ▶ No external influences on transaction trends
- ▶ High projected revenue, customer base, number of transactions and BNPL earnings, and low fraud rates lead to a high ranking merchant

RECOMMENDATIONS AND FUTURE STEPS



RECOMMENDATIONS

CONSIDER MERCHANTS WITH...



- Past high number of transactions



- Past high revenue



- Past high number of customers



- Customer base primarily located in high income per person areas



- Low average merchant fraud rate



LIMITATIONS & DIFFICULTIES



Limited internal dataset



Dataset is not reflective of real world trends



Differing device capabilities across team members

\$ WORTH PURSUING? YES!

- ▶ BNPL companies and schemes becoming more popular
- ▶ Team completed project in 6 weeks + found meaningful insights to apply to real data
- ▶ Delivering an actual product: similar time frame
 - ▶ Practical, critical, agile approach to problem
 - ▶ Apply developed algorithms to real + supporting data



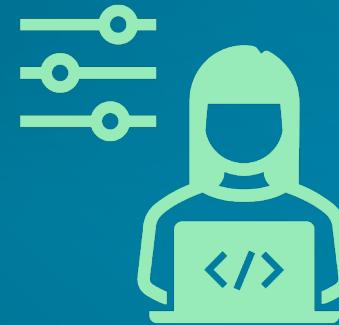
FUTURE INVESTIGATION



Real world datasets



Longer timespan



Model Adjustments



PRIVACY COMPLIANCE



Privately storing the internal data on Git Repository



Publicly available external data without API



Real people are not identifiable

THANKS!



REFERENCES

ACSC Annual Cyber Threat Report, July 2020 to June 2021 | Cyber.gov.au. (2020). Cyber.gov.au. <https://www.cyber.gov.au/acsc/view-all-content/reports-and-statistics/acsc-annual-cyber-threat-report-july-2020-june-2021>